

Optimality Bounds for a Variational Relaxation of the Image Partitioning Problem

Jan Lellmann · Frank Lenzen · Christoph Schnörr

Received: date / Accepted: date

Abstract We consider a variational convex relaxation of a class of optimal partitioning and multiclass labeling problems, which has recently proven quite successful and can be seen as a continuous analogue of Linear Programming (LP) relaxation methods for finite-dimensional problems. While for the latter case several optimality bounds are known, to our knowledge no such bounds exist in the continuous setting. We provide such a bound by analyzing a probabilistic rounding method, showing that it is possible to obtain an integral solution of the original partitioning problem from a solution of the relaxed problem with an *a priori* upper bound on the objective, ensuring the quality of the result from the viewpoint of optimization. The approach has a natural interpretation as an approximate, multiclass variant of the celebrated coarea formula.

Keywords Convex Relaxation · Multiclass Labeling · Approximation Bound · Combinatorial Optimization · Total Variation · Linear Programming Relaxation

J. Lellmann
Image and Pattern Analysis Group & HCI
Dept. of Mathematics and Computer Science, University of Heidelberg
Current Address: Dept. of Applied Mathematics and Theoretical Physics, University of Cambridge, United Kingdom
E-mail: j.ellmann@damtp.cam.ac.uk

F. Lenzen · C. Schnörr
Image and Pattern Analysis Group & HCI
Dept. of Mathematics and Computer Science
University of Heidelberg
E-mail: lenzen@iwr.uni-heidelberg.de, schnoerr@math.uni-heidelberg.de

1 Introduction and Background

1.1 Convex Relaxations of Partitioning Problems

In this work, we will be concerned with a class of *variational* problems used in image processing and analysis for formulating multiclass image partitioning problems, which are of the form

$$\inf_{u \in \mathcal{C}_{\mathcal{E}}} f(u) := \int_{\Omega} \langle u(x), s(x) \rangle dx + \int_{\Omega} d\Psi(Du), \quad (1)$$

$$\mathcal{C}_{\mathcal{E}} := \text{BV}(\Omega, \mathcal{E}) \quad (2)$$

$$= \{u \in \text{BV}(\Omega)^l \mid u(x) \in \mathcal{E} \text{ for a.e. } x \in \Omega\}, \quad (3)$$

$$\mathcal{E} := \{e^1, \dots, e^l\}. \quad (4)$$

The *labeling function* $u : \Omega \rightarrow \mathbb{R}^l$ assigns to each point in the image domain $\Omega := (0, 1)^d$ a label $i \in \mathcal{I} := \{1, \dots, l\}$, which is represented by one of the l -dimensional unit vectors e^1, \dots, e^l . Since it is piecewise constant and therefore cannot be assumed to be differentiable, the problem is formulated as a *free discontinuity problem* in the space $\text{BV}(\Omega, \mathcal{E})$ of functions of bounded variation; we refer to [2] for a general overview.

The objective function f consists of a data term and a regularizer. The data term is given in terms of the L^1 function $s(x) = (s_1(x), \dots, s_l(x)) \in \mathbb{R}^l$, and assigns to the choice $u(x) = e^i$ the “penalty” $s_i(x)$, in the sense that

$$\int_{\Omega} \langle u(x), s(x) \rangle dx = \sum_{i=1}^l \int_{\Omega_i} s_i(x) dx, \quad (5)$$

where $\Omega_i := u^{-1}(\{e^i\}) = \{x \in \Omega \mid u(x) = e^i\}$ is the *class region* for label i , i.e., the set of points that are assigned the i -th label. The data term generally depends on the input data – such as color values of a recorded image, depth measurements, or other features – and promotes a good fit of the minimizer to the input data. While it

is purely local, there are no further restrictions such as continuity, convexity etc., therefore it covers many interesting applications such as segmentation, multi-view reconstruction, stitching, and inpainting [24].

1.2 Convex Regularizers

The *regularizer* is defined by the positively homogeneous, continuous and convex function $\Psi : \mathbb{R}^{d \times l} \rightarrow \mathbb{R}_{\geq 0}$ acting on the distributional derivative Du of u , and incorporates additional prior knowledge about the “typical” appearance of the desired output. For piecewise constant u , it can be seen that the definition in (1) amounts to a weighted penalization of the *discontinuities* of u :

$$\begin{aligned} \int_{\Omega} d\Psi(Du) &= \\ &\int_{J_u} \Psi(\nu_u(x)(u^+(x) - u^-(x))^{\top}) d\mathcal{H}^{d-1}(x), \end{aligned} \quad (6)$$

where J_u is the jump set of u , i.e., the set of points where u has well-defined right-hand and left-hand limits u^+ and u^- and (in an infinitesimal sense) jumps between the values $u^+(x), u^-(x) \in \mathbb{R}^l$ across a hyperplane with normal $\nu_u(x) \in \mathbb{R}^d$, $\|\nu_u(x)\|_2 = 1$ (see [2] for the precise definitions).

A particular case is to set $\Psi = (1/\sqrt{2})\|\cdot\|_2$, i.e., the scaled Frobenius norm. In this case $J(u)$ is just the (scaled) total variation of u , and, since $u^+(x)$ and $u^-(x)$ assume values in \mathcal{E} and cannot be equal on the jump set J_u , it holds that

$$\begin{aligned} J(u) &= \frac{1}{\sqrt{2}} \int_{J_u} \|u^+(x) - u^-(x)\|_2 d\mathcal{H}^{d-1}(x), \\ &= \mathcal{H}^{d-1}(J_u). \end{aligned} \quad (7)$$

Therefore, for $\Psi = (1/\sqrt{2})\|\cdot\|_2$ the regularizer just amounts to penalizing the total length of the interfaces between class regions as measured by the $(d-1)$ -dimensional Hausdorff measure \mathcal{H}^{d-1} , which is known as *uniform metric* or *Potts* regularizer.

A general regularizer was proposed in [18], based on [5]: Given a metric (*distance*) $d : \{1, \dots, l\}^2 \rightarrow \mathbb{R}_{\geq 0}$ (not to be confused with the ambient space dimension), define

$$\Psi_d(z = (z^1, \dots, z^l)) := \sup_{v \in \mathcal{D}_{\text{loc}}^d} \langle z, v \rangle, \quad (9)$$

$$\begin{aligned} \mathcal{D}_{\text{loc}}^d &:= \{(v^1, \dots, v^l) \in \mathbb{R}^{d \times l} \mid \dots \\ &\quad \|v^i - v^j\|_2 \leq d(i, j) \forall i, j \in \{1, \dots, l\}, \dots \\ &\quad \sum_{k=1}^l v^k = 0\}. \end{aligned} \quad (10)$$

It was then shown that

$$\Psi_d(\nu(e^j - e^i)^{\top}) = d(i, j), \quad (11)$$

therefore in view of (7) the corresponding regularizer is *non-uniform*: the boundary between the class regions Ω_i and Ω_j is penalized by its length, multiplied by the weight $d(i, j)$ *depending on the labels of both regions*.

However, even for the comparably simple regularizer (7), the model (1) is a (spatially continuous) *combinatorial* problem due to the integral nature of the constraint set $\mathcal{C}_{\mathcal{E}}$, therefore optimization is nontrivial. In the context of multiclass image partitioning, a first approach can be found in [20], where the problem was posed in a level set-formulation in terms of a labeling function $\phi : \Omega \rightarrow \{1, \dots, l\}$, which is subsequently relaxed to \mathbb{R} . Then ϕ is replaced by polynomials in ϕ , which coincide with the indicator functions u_i for the case where ϕ assumes integral values. However, the numerical approach involves several nonlinearities and requires to solve a sequence of nontrivial subproblems.

In contrast, representation (1) directly suggests a more straightforward relaxation to a convex problem: replace \mathcal{E} by its convex hull, which is just the unit simplex in l dimensions,

$$\begin{aligned} \Delta_l &:= \text{conv}\{e^1, \dots, e^l\} \\ &= \{a \in \mathbb{R}^l \mid a \geq 0, \sum_{i=1}^l a_i = 1\}, \end{aligned} \quad (12)$$

and solve the *relaxed* problem

$$\inf_{u \in \mathcal{C}} f(u), \quad (13)$$

$$\mathcal{C} := \text{BV}(\Omega, \Delta_l) \quad (14)$$

$$= \{u \in \text{BV}(\Omega)^l \mid u(x) \in \Delta_l \text{ for a.e. } x \in \Omega\}. \quad (15)$$

Sparked by a series of papers [29, 5, 17], recently there has been much interest in problems of this form, since they – although generally nonsmooth – are convex and therefore can be solved to global optimality, e.g., using primal-dual techniques. The approach has proven useful for a wide range of applications [14, 11, 10, 28].

1.3 Finite-Dimensional vs. Continuous Approaches

Many of these applications have been tackled before in a finite-dimensional setting, where they can be formulated as combinatorial problems on a grid graph, and solved using combinatorial optimization methods such as α -expansion and related integer linear programming (ILP) methods [4, 15]. These methods have been shown to yield an integral labeling $u' \in \mathcal{C}_{\mathcal{E}}$ with the a priori bound

$$f(u') \leq 2 \frac{\max_{i \neq j} d(i, j)}{\min_{i \neq j} d(i, j)} f(u_{\mathcal{E}}^*), \quad (16)$$

where $u_{\mathcal{E}}^*$ is the (unknown) solution of the integral problem (1). They therefore permit to compute a suboptimal solution to the – originally NP-hard [4] – combinatorial problem with an *upper bound* on the objective.

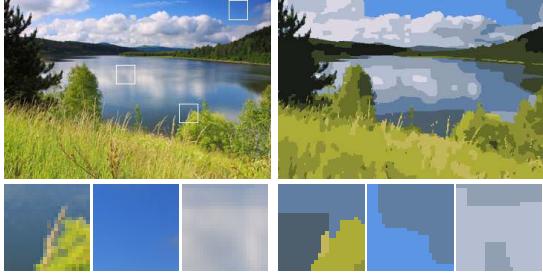


Figure 1 Segmentation of an image into 12 classes using a combinatorial method. **Left:** Input image, **Right:** Result obtained by solving a *combinatorial* discretized problem with 4-neighborhood. The bottom row shows detailed views of the marked parts of the image. The minimizer of the combinatorial problem exhibits blocky artifacts caused by the choice of discretization.

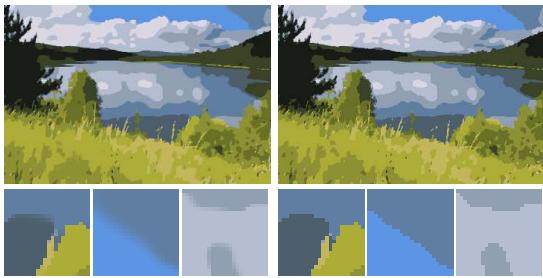


Figure 2 Segmentation obtained by solving a finite-differences discretization of the *relaxed* spatially continuous problem. **Left:** Non-integral solution obtained as a minimizer of the discretized relaxed problem. **Right:** Integral labeling obtained by rounding the fractional labels in the solution of the relaxed problem to the nearest integral label. The rounded result contains almost no structural artifacts.

No such bound is yet available for methods based on the spatially continuous problem (13).

Despite these strong theoretical and practical results available for the finite-dimensional combinatorial energies, the function-based, spatially continuous formulation (1) has several unique advantages:

- The energy (1) is truly isotropic, in the sense that for a proper choice of Ψ it is invariant under rotation of the coordinate system. Pursuing finite-dimensional “discretize-first” approaches generally introduces artifacts due to the inherent anisotropy, which can only be avoided by increasing the neighborhood size, thereby reducing sparsity and severely slowing down the graph cut-based methods.

In contrast, properly discretizing the *relaxed* problem (13) and solving it as a *convex* problem with subsequent thresholding yields much better results without compromising sparsity (Fig. 1 and 2, [13]). This can be attributed to the fact that solving the discretized problem as a *combinatorial* problem in effect discards much of the information about the

problem structure that is contained in the nonlinear terms of the discretized objective.

- Present combinatorial optimization methods [4, 15] are inherently sequential and difficult to parallelize. On the other hand, parallelizing primal-dual methods for solving the relaxed problem (13) is straightforward, and GPU implementations have been shown to outperform state-of-the-art graph cut methods [29].
- Analyzing the problem in a fully functional-analytic setting gives valuable insight into the problem structure, and is of theoretical interest in itself.

1.4 Optimality Bounds

However, one possible drawback of the spatially continuous approach is that the solution of the relaxed problem (13) may assume *fractional* values, i.e., values in $\Delta_l \setminus \mathcal{E}$. Therefore, in applications that require a true partition of Ω , some rounding process is needed in order to generate an integral labeling \bar{u}^* . This may increase the objective, and lead to a suboptimal solution of the original problem (1).

The regularizer Ψ_d as defined in (9) enjoys the property that it majorizes all other regularizers that can be written in integral form and satisfy (11). Therefore it is in a sense “optimal”, since it introduces as few fractional solutions as possible. In practice, this forces solutions of the relaxed problem to assume integral values in most points, and rounding is in practice only required in small regions.

However, the rounding step may still increase the objective and generate suboptimal integral solutions. Therefore the question arises whether this approach allows to recover “good” integral solutions of the original problem (1).

In the following, we are concerned with the question whether it is possible to obtain, using the convex relaxation (13), *integral* solutions with an upper bound on the objective. Specifically, we focus on inequalities of the form

$$f(\bar{u}^*) \leq (1 + \varepsilon) f(u_{\mathcal{E}}^*) \quad (17)$$

for some constant $\varepsilon > 0$, which provide an *upper bound on the objective* of the *rounded* integral solution \bar{u}^* with respect to the objective of the (unknown) *optimal* integral solution $u_{\mathcal{E}}^*$ of (1). Note that generally it is not possible to show that (17) holds for *any* $\varepsilon > 0$. The reverse inequality

$$f(u_{\mathcal{E}}^*) \leq f(\bar{u}^*) \quad (18)$$

always holds since $\bar{u}^* \in \mathcal{C}_{\mathcal{E}}$ and $u_{\mathcal{E}}^*$ is an optimal integral solution. The specific form (17) can be attributed to the

alternative interpretation

$$\frac{f(\bar{u}^*) - f(u_{\mathcal{E}}^*)}{f(u_{\mathcal{E}}^*)} \leq \varepsilon, \quad (19)$$

which provides a bound for the relative gap to the optimal objective of the combinatorial problem. Such ε can be obtained *a posteriori* by actually computing (or approximating) \bar{u}^* and a dual feasible point: Assume that a feasible primal-dual pair $(u, v) \in \mathcal{C} \times \mathcal{D}$ is known, where u approximates u^* , and assume that some integral feasible $\bar{u} \in \mathcal{C}_{\mathcal{E}}$ has been obtained from u by a rounding process. Then the pair (\bar{u}, v) is feasible as well since $\mathcal{C}_{\mathcal{E}} \subseteq \mathcal{C}$, and we obtain an *a posteriori* optimality bound of the form (19) with respect to the optimal *integral* solution $u_{\mathcal{E}}^*$:

$$\frac{f(\bar{u}) - f_D(u_{\mathcal{E}}^*)}{f_D(u_{\mathcal{E}}^*)} \leq \frac{f(\bar{u}) - f_D(u^*)}{f_D(v)} \leq \frac{f(\bar{u}) - f_D(v)}{f_D(v)} =: \varepsilon'. \quad (20)$$

However, this requires that the the primal and dual objectives f and f_D can be accurately evaluated, and requires to compute a minimizer of the problem for the specific input data, which is generally difficult, especially in the spatially continuous formulation.

In contrast, true *a priori* bounds do not require knowledge of a solution and apply uniformly to all problems of a class, irrespective of the particular input. When considering rounding methods, one generally has to discriminate between

- *deterministic* vs. *probabilistic* methods, and
- *spatially discrete (finite-dimensional)* vs. *spatially continuous* methods.

Most known *a priori* approximation results only hold in the finite-dimensional setting, and are usually proven using graph-based pairwise formulations. In contrast, we assume an “optimize first” perspective due to the reasons outlined in the introduction. Unfortunately, the proofs for the finite-dimensional results often rely on pointwise arguments that cannot directly be transferred to the continuous setting. Deriving similar results for continuous problems therefore requires considerable additional work.

1.5 Contribution and Main Results

In this work we prove that using the regularizer (9), the *a priori* bound (16) can be carried over to the spatially continuous setting. Preliminary versions of these results with excerpts of the proofs have been announced as conference proceedings [18]. We extend these results to provide the exact bound (16), and supply the full proofs.

As the main result, we show that it is possible to construct a rounding method parametrized by a parameter $\gamma \in \Gamma$, where Γ is an appropriate parameter space:

$$R : \mathcal{C} \times \Gamma \rightarrow \mathcal{C}_{\mathcal{E}}, \quad (21)$$

$$u \in \mathcal{C} \mapsto \bar{u}_{\gamma} := R_{\gamma}(u) \in \mathcal{C}_{\mathcal{E}}, \quad (22)$$

such that for a suitable probability distribution on Γ , the following theorem holds for the expectation $\mathbb{E}f(\bar{u}) := \mathbb{E}_{\gamma}f(\bar{u}_{\gamma})$:

Theorem 1 *Let $u \in \mathcal{C}$, $s \in L^1(\Omega)^l$, $s \geq 0$, and let $\Psi : \mathbb{R}^{d \times l} \rightarrow \mathbb{R}_{\geq 0}$ be positively homogeneous, convex and continuous. Assume there exists a lower bound $\lambda_l > 0$ such that, for $z = (z^1, \dots, z^l)$,*

$$\Psi(z) \geq \lambda_l \frac{1}{2} \sum_{i=1}^l \|z^i\|_2 \quad \forall z \in \mathbb{R}^{d \times l}, \sum_{i=1}^l z^i = 0. \quad (23)$$

Moreover, assume there exists an upper bound $\lambda_u < \infty$ such that, for every $\nu \in \mathbb{R}^d$ satisfying $\|\nu\|_2 = 1$,

$$\Psi(\nu(e^i - e^j)^{\top}) \leq \lambda_u \quad \forall i, j \in \{1, \dots, l\}. \quad (24)$$

Then Alg. 1 (see below) generates an integral labeling $\bar{u} \in \mathcal{C}_{\mathcal{E}}$ almost surely, and

$$\mathbb{E}f(\bar{u}) \leq 2 \frac{\lambda_u}{\lambda_l} f(u). \quad (25)$$

Note that $\lambda_u \geq \lambda_l$ always holds if both are defined, since (23) implies, for ν with $\|\nu\|_2 = 1$,

$$\lambda_u \geq \Psi(\nu(e^i - e^j)^{\top}) \geq \frac{\lambda_l}{2} (\|\nu\|_2 + \|\nu\|_2) = \lambda_l. \quad (26)$$

The proof of Thm. 1 (Sect. 4) is based on the work of Kleinberg and Tardos [12], which is set in an LP relaxation framework. However their results are restricted in that they assume a graph-based representation and extensively rely on the finite dimensionality. In contrast, our results hold in the continuous setting without assuming a particular problem discretization.

Theorem 1 guarantees that – in a probabilistic sense – the rounding process may only increase the energy in a controlled way, with an upper bound depending on Ψ . An immediate consequence is

Corollary 1 *Under the conditions of Thm. 1, if u^* minimizes f over \mathcal{C} , $u_{\mathcal{E}}^*$ minimizes f over $\mathcal{C}_{\mathcal{E}}$, and \bar{u}^* denotes the output of Alg. 1 applied to u^* , then*

$$\mathbb{E}f(\bar{u}^*) \leq 2 \frac{\lambda_u}{\lambda_l} f(u_{\mathcal{E}}^*). \quad (27)$$

Therefore the proposed approach allows to recover, from the solution u^* of the convex *relaxed* problem (13), an approximate *integral* solution \bar{u}^* of the nonconvex

original problem (1) with an upper bound on the objective.

In particular, for the tight relaxation of the regularizer as in (9), we obtain (cf. Prop. 10)

$$\mathbb{E}f(\bar{u}^*) \leq 2 \frac{\lambda_u}{\lambda_l} = 2 \frac{\max_{i \neq j} d(i, j)}{\min_{i \neq j} d(i, j)}, \quad (28)$$

which is exactly the same bound as has been proven for the combinatorial α -expansion method (16).

To our knowledge, this is the first bound available for the fully spatially convex relaxed problem (13). Related is the work of Olsson et al. [22, 23], where the authors consider a continuous analogue to the α -expansion method known as continuous binary fusion [27], and claim that a bound similar to (16) holds for the corresponding fixed points when using the separable regularizer

$$\Psi_A(z) := \sum_{j=1}^l \|Az^j\|_2, \quad z \in \mathbb{R}^{d \times l}, \quad (29)$$

for some $A \in \mathbb{R}^{d \times d}$, which implements an anisotropic variant of the uniform metric. However, a rigorous proof in the BV framework was not given.

In [3], the authors propose to solve the problem (1) by considering the dual problem to (13) consisting of l coupled maximum-flow problems, which are solved using a log-sum-exp smoothing technique and gradient descent. In case the dual solution allows to unambiguously recover an (integral) primal solution, the latter is necessarily the unique minimizer of f , and therefore a global *integral* minimizer of the combinatorial problem (1). This provides an *a posteriori* bound, which applies if a dual solution can be computed. While useful in practice as a certificate for global optimality, in the spatially continuous setting it requires explicit knowledge of a dual solution, which is rarely available since it depends on the regularizer Ψ as well as the input data s .

In contrast, the *a priori* bound (27) holds uniformly over all problem instances, does not require knowledge of any primal or dual solutions and covers also non-uniform regularizers.

2 A Probabilistic View of the Coarea Formula

2.1 The Two-Class Case

As a motivation for the following sections, we first provide a probabilistic interpretation of a tool often used in geometric measure theory, the *coarea formula* (cf. [2]). Assuming $u' \in \text{BV}(\Omega)$ and $u'(x) \in [0, 1]$ for a.e. $x \in \Omega$, the coarea formula states that the total variation of u'

can be represented by summing the boundary lengths of its super-levelsets:

$$\text{TV}(u') = \int_0^1 \text{TV}(1_{\{u' > \alpha\}}) d\alpha. \quad (30)$$

Here 1_A denotes the characteristic function of a set A , i.e., $1_A(x) = 1$ iff $x \in A$ and $1_A(x) = 0$ otherwise. The coarea formula provides a connection between problem (1) and the relaxation (13) in the two-class case, where $\mathcal{E} = \{e^1, e^2\}$, $u \in \mathcal{C}_{\mathcal{E}}$ and $u_1 = 1 - u_2$: As noted in [16],

$$\text{TV}(u) = \|e^1 - e^2\|_2 \text{TV}(u_1) = \sqrt{2} \text{TV}(u_1), \quad (31)$$

therefore the coarea formula (30) can be rewritten as

$$\text{TV}(u) = \sqrt{2} \int_0^1 \text{TV}(1_{\{u_1 > \alpha\}}) d\alpha \quad (32)$$

$$= \int_0^1 \text{TV}(e^1 1_{\{u_1 > \alpha\}} + e^2 1_{\{u_1 \leq \alpha\}}) d\alpha \quad (33)$$

$$= \int_0^1 \text{TV}(\bar{u}_{\alpha}) d\alpha, \quad (34)$$

$$\bar{u}_{\alpha} := e^1 1_{\{u_1 > \alpha\}} + e^2 1_{\{u_1 \leq \alpha\}}. \quad (35)$$

Consequently, the total variation of u can be expressed as the *mean* over the total variations of a set of *integral* labelings $\{\bar{u}_{\alpha} \in \mathcal{C}_{\mathcal{E}} | \alpha \in [0, 1]\}$, obtained by *rounding* u at different thresholds α . We now adopt a *probabilistic* view of (35): We regard the mapping

$$R : (u, \alpha) \in \mathcal{C} \times [0, 1] \mapsto \bar{u}_{\alpha} \in \mathcal{C}_{\mathcal{E}} \quad (\text{a.e. } \alpha \in [0, 1]) \quad (36)$$

as a *parametrized, deterministic* rounding algorithm that depends on u and on an additional parameter α . From this we obtain a *probabilistic* (randomized) rounding algorithm by assuming α to be a uniformly distributed random variable. Under these assumptions the coarea formula (35) can be written as

$$\text{TV}(u) = \mathbb{E}_{\alpha} \text{TV}(\bar{u}_{\alpha}). \quad (37)$$

This has the probabilistic interpretation that applying the probabilistic rounding to (arbitrary, but fixed) u does – in a probabilistic sense, i.e., in the mean – not change the objective. It can be shown that this property extends to the full functional f in (13): In the two-class case, the “coarea-like” property

$$f(u) = \mathbb{E}_{\alpha} f(\bar{u}_{\alpha}) \quad (38)$$

holds. Functions with property (38) are also known as *levelable functions* [8, 9] or *discrete total variations* [6] and have been studied in [26]. A well-known implication is that if $u = u^*$, i.e., u minimizes the relaxed problem (13), then in the two-class case almost every $\bar{u}^* = \bar{u}_{\alpha}^*$ is an integral minimizer of the original problem (1), i.e., the optimality bound (17) holds with $\varepsilon = 0$ [7].

2.2 The Multi-Class Case and Generalized Coarea Formulas

Generalizing these observations to more than two labels hinges on a property similar to (38) that holds for *vector-valued* u . In a general setting, the question is whether there exist

- a probability space (Γ, μ) , and
- a *parametrized rounding method*, i.e., for μ -almost every $\gamma \in \Gamma$:

$$R : \mathcal{C} \times \Gamma \rightarrow \mathcal{C}_{\mathcal{E}}, \quad (39)$$

$$u \in \mathcal{C} \mapsto \bar{u}_{\gamma} := R_{\gamma}(u) \in \mathcal{C}_{\mathcal{E}} \quad (40)$$

satisfying $R_{\gamma}(u') = u'$ for all $u' \in \mathcal{C}_{\mathcal{E}}$,

such that a “multiclass coarea-like property” (or *generalized coarea formula*)

$$f(u) = \int_{\Gamma} f(\bar{u}_{\gamma}) d\mu(\gamma) \quad (41)$$

holds. In a probabilistic sense this corresponds to

$$f(u) = \int_{\Gamma} f(\bar{u}_{\gamma}) d\mu(\gamma) = \mathbb{E}_{\gamma} f(\bar{u}_{\gamma}). \quad (42)$$

For $l = 2$ and $\Psi(x) = \|\cdot\|_2$, (37) shows that (42) holds with $\gamma = \alpha$, $\Gamma = [0, 1]$, $\mu = \mathcal{L}^1$, and $R : \mathcal{C} \times \Gamma \rightarrow \mathcal{C}_{\mathcal{E}}$ as defined in (36). Unfortunately, property (37) is intrinsically restricted to the two-class case with TV regularizer.

In the multiclass case, the difficulty lies in providing a suitable combination of a probability space (Γ, μ) and a parametrized rounding step $(u, \gamma) \mapsto \bar{u}_{\gamma}$. Unfortunately, obtaining a relation such as (37) for the full functional (1) is unlikely, as it would mean that solutions to the (after discretization) NP-hard problem (1) could be obtained by solving the convex relaxation (13) and subsequent rounding, which can be achieved in polynomial time.

Therefore we restrict ourselves to an *approximate* variant of the generalized coarea formula:

$$(1 + \varepsilon)f(u) \geq \int_{\Gamma} f(\bar{u}_{\gamma}) d\mu(\gamma) = \mathbb{E}_{\gamma} f(\bar{u}_{\gamma}). \quad (43)$$

While (43) is not sufficient to provide a bound on $f(\bar{u}_{\gamma})$ for *particular* γ , it permits a *probabilistic* bound: for any minimizer u^* of the relaxed problem (13), eq. (43) implies

$$\mathbb{E}_{\gamma} f(\bar{u}_{\gamma}^*) \leq (1 + \varepsilon)f(u^*) \leq (1 + \varepsilon)f(u_{\mathcal{E}}^*), \quad (44)$$

i.e., the ratio between the objective of the *rounded relaxed solution* and the *optimal integral solution* is bounded – in a probabilistic sense – by $(1 + \varepsilon)$.

In the following sections we construct a suitable parametrized rounding method and probability space in order to obtain an approximate generalized coarea formula of the form (43).

Algorithm 1 Continuous Probabilistic Rounding

```

1:  $u^0 \leftarrow u$ ,  $U^0 \leftarrow \Omega$ ,  $c^0 \leftarrow (1, \dots, 1) \in \mathbb{R}^l$ .
2: for  $k = 1, 2, \dots$  do
3:   Randomly choose  $\gamma^k = (i^k, \alpha^k) \in \mathcal{I} \times [0, 1]$  uniformly.
4:    $M^k \leftarrow U^{k-1} \cap \{x \in \Omega | u_{i^k}^{k-1}(x) > \alpha^k\}$ .
5:    $u^k \leftarrow e^{i^k} 1_{M^k} + u^{k-1} 1_{\Omega \setminus M^k}$ .
6:    $U^k \leftarrow U^{k-1} \setminus M^k$ .
7:    $c_j^k \leftarrow \begin{cases} \min\{c_j^{k-1}, \alpha^k\}, & j = i^k, \\ c_j^{k-1}, & \text{otherwise.} \end{cases}$ 
8: end for

```

3 Probabilistic Rounding for Multiclass Image Partitions

3.1 Approach

We consider the probabilistic rounding approach based on [12] as defined in Alg. 1.

The algorithm proceeds in a number of phases. At each iteration, a label and a threshold

$$\gamma^k := (i^k, \alpha^k) \in \Gamma' := \mathcal{I} \times [0, 1]$$

are randomly chosen (step 3), and label i^k is assigned to all yet unassigned points x where $u_{i^k}^{k-1}(x) > \alpha^k$ holds (step 5). In contrast to the two-class case considered above, the randomness is provided by a *sequence* (γ^k) of uniformly distributed random variables, i.e., $\Gamma = (\Gamma')^{\mathbb{N}}$.

After iteration k , all points in the set $U^k \subseteq \Omega$ are still *unassigned*, while all points in $\Omega \setminus U^k$ have been assigned an (integral) label in iteration k or in a previous iteration. Iteration $k+1$ potentially modifies points only in the set U^k . The variable c_j^k stores the lowest threshold α chosen for label j up to and including iteration k , and is only required for the proofs.

While the algorithm is defined using pointwise operations, it is well-defined in the sense that for fixed γ , the sequence (u^k) , viewed as elements in L^1 , does not depend on the specific representative of u in its equivalence class in L^1 . The sequences (M^k) and (U^k) depend on the representative, but are unique up to \mathcal{L}^d -negligible sets.

In an actual implementation, the algorithm could be terminated as soon as all points in Ω have been assigned a label, i.e., $U^k = \emptyset$. However, in our framework used for analysis the algorithm never terminates explicitly. Instead, for fixed input u we regard the algorithm as a mapping between *sequences* of parameters (or instances of random variables) $\gamma = (\gamma^k) \in \Gamma$ and *sequences* of states (u_{γ}^k) , (U_{γ}^k) and (c_{γ}^k) . We drop the subscript γ if it does not create ambiguities. The elements of the sequence (γ^k) are independently uniformly distributed, and by the Kolmogorov extension theorem [21, Thm. 2.1.5] there exists a probability space and a stochas-

tic process on the set of sequences γ with compatible marginal distributions.

In order to define the parametrized rounding step $(u, \gamma) \mapsto \bar{u}_\gamma$, we observe that once $U_\gamma^{k'} = \emptyset$ occurs for some $k' \in \mathbb{N}$, the sequence (u_γ^k) becomes stationary at $u_\gamma^{k'}$. In this case the algorithm may be terminated, with output $\bar{u}_\gamma := u_\gamma^{k'}$:

Definition 1 Let $u \in \text{BV}(\Omega)^l$ and $f : \text{BV}(\Omega)^l \rightarrow \mathbb{R}$. For some $\gamma \in \Gamma$, if $U_\gamma^{k'} = \emptyset$ in Alg. 1 for some $k' \in \mathbb{N}$, we denote $\bar{u}_\gamma := u_\gamma^{k'}$. We define

$$f(\bar{u}_{(\cdot)}) : \Gamma \rightarrow \mathbb{R} \cup \{+\infty\}, \gamma \in \Gamma \mapsto f(\bar{u}_\gamma), \quad (45)$$

$$f(\bar{u}_\gamma) := \begin{cases} f(u_\gamma^{k'}), \exists k' \in \mathbb{N} : U_\gamma^{k'} = \emptyset \wedge u_\gamma^{k'} \in \text{BV}(\Omega)^l, \\ +\infty, \text{ otherwise.} \end{cases}$$

We denote by $f(\bar{u})$ the corresponding random variable induced by assuming γ to be uniformly distributed on Γ .

As indicated above, $f(\bar{u}_\gamma)$ is well-defined: if $U_\gamma^{k'} = \emptyset$ for some (γ, k') then $u_\gamma^{k'} = u_\gamma^{k''}$ for all $k'' \geq k'$. Instead of focusing on local properties of the random sequence (u_γ^k) as in the proofs for the finite-dimensional case, we derive our results directly for the sequence $(f(u_\gamma^k))$. In particular, we show that the expectation of $f(\bar{u})$ over all sequences γ can be bounded according to

$$\mathbb{E}f(\bar{u}) = \mathbb{E}_\gamma f(\bar{u}_\gamma) \leq (1 + \varepsilon)f(\bar{u}) \quad (46)$$

for some $\varepsilon \geq 0$, cf. (43). Consequently, the rounding process may only increase the average objective in a controlled way.

3.2 Termination Properties

Theoretically, the algorithm may produce a sequence (u_γ^k) that does *not* become stationary, or becomes stationary with a solution that is not an element of $\text{BV}(\Omega)^l$. In Thm. 2 below we show that this happens only with zero probability, i.e., almost surely Alg. 1 generates (in a finite number of iterations) an *integral* labeling function $\bar{u}_\gamma \in \mathcal{C}_\mathcal{E}$. The following two propositions are required for the proof.

Proposition 1 For the sequence (c^k) generated by Algorithm 1,

$$\mathbb{P}(e^\top c^k < 1) \geq \sum_{p \in \{0,1\}^l} (-1)^{e^\top p} \left(\sum_{j=1}^l \frac{1}{l} \left(\left(1 - \frac{1}{l}\right)^{p_j} \right) \right)^k \quad (47)$$

holds. In particular,

$$\mathbb{P}(e^\top c^k < 1) \xrightarrow{k \rightarrow \infty} 1. \quad (48)$$

Proof Denote by $n_j^k \in \mathbb{N}_0$ the number of $k' \in \{1, \dots, k\}$ such that $i^{k'} = j$, i.e., the number of times label j was selected up to and including the k -th step. Then

$$(n_1^k, \dots, n_l^k) \sim \text{Multinomial}\left(k; \frac{1}{l}, \dots, \frac{1}{l}\right), \quad (49)$$

i.e., the probability of a specific instance is

$$\mathbb{P}((n_1^k, \dots, n_l^k)) = \begin{cases} \frac{k!}{n_1^k! \cdots n_l^k!} \left(\frac{1}{l}\right)^k, & \sum_j n_j^k = k, \\ 0, & \text{otherwise.} \end{cases} \quad (50)$$

Therefore,

$$\mathbb{P}(e^\top c^k < 1) = \sum_{n_1^k, \dots, n_l^k} \mathbb{P}(e^\top c^k < 1 | (n_1^k, \dots, n_l^k)) \cdots \mathbb{P}((n_1^k, \dots, n_l^k)) \quad (51)$$

$$= \sum_{n_1^k + \dots + n_l^k = k} \frac{k!}{n_1^k! \cdots n_l^k!} \left(\frac{1}{l}\right)^k \cdots \mathbb{P}(e^\top c^k < 1 | (n_1^k, \dots, n_l^k)). \quad (52)$$

Since $c_1^k, \dots, c_l^k < \frac{1}{l}$ is a sufficient condition for $e^\top c < 1$, we may bound the probability according to

$$\mathbb{P}(e^\top c < 1) \geq \sum_{n_1^k + \dots + n_l^k = k} \frac{k!}{n_1^k! \cdots n_l^k!} \left(\frac{1}{l}\right)^k \cdots \mathbb{P}\left(c_j^k < \frac{1}{l} \forall j \in \mathcal{I} | (n_1^k, \dots, n_l^k)\right). \quad (53)$$

We now consider the distributions of the components c_j^k of c^k conditioned on the vector (n_1^k, \dots, n_l^k) . Given n_j^k , the probability of $\{c_j^k \geq t\}$ is the probability that in each of the n_j^k steps where label j was selected the threshold α was randomly chosen to be *at least as large as* t . For $0 < t < 1$, we conclude

$$\mathbb{P}(c_j^k < t | (n_1^k, \dots, n_l^k)) = \mathbb{P}(c_j^k < t | n_j^k) \quad (54)$$

$$= 1 - \mathbb{P}(c_j^k \geq t | n_j^k) \quad (55)$$

$$\stackrel{0 < t < 1}{=} 1 - (1 - t)^{n_j^k}. \quad (56)$$

The above formulation also covers the case $n_j^k = 0$ (note that we assumed $0 < t < 1$). For fixed k the distributions of the c_j^k are independent when conditioned on (n_1^k, \dots, n_l^k) . Therefore we obtain from (53) and (56)

$$\mathbb{P}(e^\top c < 1) \stackrel{(53)}{\geq} \sum_{n_1^k + \dots + n_l^k = k} \frac{k!}{n_1^k! \cdots n_l^k!} \left(\frac{1}{l}\right)^k \cdots \prod_{j=1}^l \mathbb{P}\left(c_j^k < \frac{1}{l} | (n_1^k, \dots, n_l^k)\right) \quad (57)$$

$$\stackrel{(56)}{=} \sum_{n_1^k + \dots + n_l^k = k} \frac{k!}{n_1^k! \cdots n_l^k!} \left(\frac{1}{l}\right)^k + \prod_{j=1}^l \left(1 - \left(1 - \frac{1}{l}\right)^{n_j^k}\right). \quad (58)$$

Expanding the product and swapping the summation order, we derive

$$\begin{aligned} & \mathbb{P}(e^\top c^k < 1) \\ & \geq \sum_{n_1^k + \dots + n_l^k = k} \frac{k!}{n_1^k! \cdot \dots \cdot n_l^k!} \left(\frac{1}{l} \right)^k. \end{aligned} \quad (59)$$

$$\sum_{p \in \{0,1\}^l} \prod_{j=1}^l \left(- \left(1 - \frac{1}{l} \right)^{n_j^k} \right)^{p_j} \quad (60)$$

$$\begin{aligned} & = \sum_{p \in \{0,1\}^l} (-1)^{e^\top p} \sum_{n_1^k + \dots + n_l^k = k} \frac{k!}{n_1^k! \cdot \dots \cdot n_l^k!} \cdot \\ & \quad \prod_{j=1}^l \left(\frac{1}{l} \left(1 - \frac{1}{l} \right)^{p_j} \right)^{n_j^k}. \end{aligned} \quad (61)$$

Using the multinomial summation formula, we conclude

$$\begin{aligned} & \mathbb{P}(e^\top c^k < 1) \geq \\ & \quad \sum_{p \in \{0,1\}^l} (-1)^{e^\top p} \left(\underbrace{\sum_{j=1}^l \frac{1}{l} \left(1 - \frac{1}{l} \right)^{p_j}}_{=: q_p} \right)^k, \end{aligned} \quad (62)$$

which proves (47). At (*) the multinomial summation formula was invoked. Note that in (62) the n_j^k do not occur explicitly anymore. To show the second assertion (48), we use the fact that, for any $p \neq (0, \dots, 0)$, q_p can be bounded by $0 < q_p < 1$. Therefore

$$\mathbb{P}(e^\top c^k < 1) \geq q_0 + \sum_{p \in \{0,1\}^l, p \neq 0} (-1)^{e^\top p} (q_p)^k \quad (63)$$

$$= 1 + \sum_{p \in \{0,1\}^l, p \neq 0} (-1)^{e^\top p} \underbrace{(q_p)^k}_{\substack{k \rightarrow \infty \\ 0}} \quad (64)$$

$$\xrightarrow{k \rightarrow \infty} 1, \quad (65)$$

which proves (48).

We now show that Alg. 1 generates a sequence in $\text{BV}(\Omega)^l$ almost surely. The *perimeter* of a set A is defined as the total variation of its characteristic function $\text{Per}(A) := \text{TV}(1_A)$.

Proposition 2 *For the sequences (u^k) , (U^k) generated by Alg. 1, define*

$$A := \bigcap_{k=1}^{\infty} \{ \gamma \in \Gamma \mid \text{Per}(U_\gamma^k) < \infty \}. \quad (66)$$

Then

$$\mathbb{P}(A) = 1. \quad (67)$$

If $\text{Per}(U_\gamma^k) < \infty$ for all k , then $u_\gamma^k \in \text{BV}(\Omega)^l$ for all k as well. Moreover,

$$\mathbb{P}(u^k \in \text{BV}(\Omega)^l \wedge \text{Per}(U^k) < \infty \forall k \in \mathbb{N}) = 1, \quad (68)$$

i.e., the algorithm almost surely generates a sequence of BV functions (u^k) and a sequence of sets of finite perimeter (U^k) .

Proof We first show that if $\text{Per}(U^{k'}) < \infty$ for all $k' \leq k$, then $u^k \in \text{BV}(\Omega)^l$ for all $k' \leq k$ as well. For $k = 0$, the assertion holds, since $u^0 = u \in \text{BV}(\Omega)^l$ by assumption. For $k \geq 1$,

$$u^k = e^{i^k} 1_{M^k} + u^{k-1} 1_{\Omega \setminus M^k}. \quad (69)$$

Since $M^k = U^{k-1} \cap (\Omega \setminus U^k)$, and U^k, U^{k-1} are assumed to have finite perimeter, M^k also has finite perimeter. Applying [2, Thm. 3.84] together with the boundedness of u^{k-1} and $u^{k-1} \in \text{BV}(\Omega)^l$ by induction then provides $u^k \in \text{BV}(\Omega)^l$.

We now denote

$$I^k := \{ \gamma \in \Gamma \mid \text{Per}(U_\gamma^k) = \infty \}, \quad (70)$$

and the event that the *first* set with non-finite perimeter is encountered at step $k \in \mathbb{N}_0$ by

$$B^k := I^k \cap (\Gamma \setminus I^{k-1}) \cap \dots \cap (\Gamma \setminus I^0). \quad (71)$$

Then

$$\mathbb{P}(A) = 1 - \mathbb{P} \left(\bigcup_{k=0}^{\infty} B^k \right). \quad (72)$$

As the sets B^k are pairwise disjoint, and due to the countable additivity of the probability measure, we have

$$\mathbb{P}(A) = 1 - \sum_{k=0}^{\infty} \mathbb{P}(B^k). \quad (73)$$

Now $U^0 = \Omega$, therefore $\text{Per}(U^0) = \text{TV}(1_{U^0}) = 0 < \infty$ and $\mathbb{P}(B^0) = 0$. For $k \geq 1$, we have

$$\begin{aligned} \mathbb{P}(B^k) & \leq \mathbb{P} \left(\text{Per}(U^k) = \infty \wedge \text{Per}(U^{k'}) < \infty \ \forall k' < k \right) \\ & \leq \mathbb{P} \left(\text{Per}(U^k) = \infty \mid \text{Per}(U^{k'}) < \infty \ \forall k' < k \right) \\ & = \mathbb{P} \left(\text{Per}(U^{k-1} \cap \{x \in \Omega \mid u_{i^k}^{k-1}(x) \leq \alpha^k\}) = \infty \mid \text{Per}(U^{k'}) < \infty \ \forall k' < k \right). \end{aligned} \quad (74)$$

By the argument from the beginning of the proof, we know that $u^{k-1} \in \text{BV}(\Omega)^l$ under the condition on the perimeter $\text{Per}(U^{k'})$, therefore from [2, Thm. 3.40] we conclude that $\text{Per}(\{x \in \Omega \mid u_{i^k}^{k-1}(x) \leq \alpha^k\})$ is finite for \mathcal{L}^1 -a.e. α^k and all i^k . As the sets of finite perimeter are closed under finite intersection, and since the α^k are drawn from an uniform distribution, this implies that

$$\mathbb{P}(\text{Per}(U^k) < \infty \mid \text{Per}(U^{k-1}) < \infty) = 1. \quad (75)$$

Together with (74) we arrive at

$$\mathbb{P}(B^k) = 0. \quad (76)$$

Substituting this result into (73) leads to the assertion, $\mathbb{P}(A) = 1$.

Equation (68) follows immediately.

Using these propositions, we now formulate the main result of this section: Alg. 1 almost surely generates an integral labeling that is of bounded variation.

Theorem 2 *Let $u \in \text{BV}(\Omega)^l$ and $f(\bar{u})$ as in Def. 1. Then*

$$\mathbb{P}(f(\bar{u}) < \infty) = 1. \quad (78)$$

Proof The first part is to show that (u^k) becomes stationary almost surely, i.e.,

$$\mathbb{P}(\exists k \in \mathbb{N} : U^k = \emptyset) = 1. \quad (79)$$

Assume there exists k such that $e^\top c^k < 1$, and assume further that $U^k \neq \emptyset$, i.e., there exists some $x \in U^k$. Then $u_j(x) \leq c_j^k$ for all labels j . But then $e^\top u(x) \leq e^\top c^k < 1$, which is a contradiction to $u(x) \in \Delta_l$. Therefore U^k must be empty. From this observation and Prop. 1 we conclude, for all $k' \in \mathbb{N}$,

$$1 \geq \mathbb{P}(\exists k \in \mathbb{N} : U^k = \emptyset) \geq \mathbb{P}(e^\top c^{k'} < 1) \xrightarrow{k' \rightarrow \infty} 1, \quad (80)$$

which proves (79).

In order to show that $f(\bar{u}_\gamma) < \infty$ with probability 1, it remains to show that the result is almost surely in $\text{BV}(\Omega)^l$. A sufficient condition is that almost surely all iterates u^k are elements of $\text{BV}(\Omega)^l$, i.e.,

$$\mathbb{P}(u^k \in \text{BV}(\Omega)^l \quad \forall k \in \mathbb{N}) = 1. \quad (81)$$

This is shown by Prop. 2. Then

$$\mathbb{P}(f(\bar{u}) < \infty) \quad (82)$$

$$\geq \mathbb{P}(\{\exists k \in \mathbb{N} : U^k = \emptyset\} \wedge \{u^k \in \text{BV}(\Omega)^l \quad \forall k \in \mathbb{N}\})$$

$$= \mathbb{P}(\{u^k \in \text{BV}(\Omega)^l \quad \forall k \in \mathbb{N}\}) \quad (83)$$

$$- \mathbb{P}(\{\forall k \in \mathbb{N} : U^k \neq \emptyset\} \wedge \{u^k \in \text{BV}(\Omega)^l \quad \forall k \in \mathbb{N}\})$$

$$\stackrel{(81)}{=} \mathbb{P}(\{u^k \in \text{BV}(\Omega)^l \quad \forall k \in \mathbb{N}\}) - 0 \quad (84)$$

$$= 1. \quad (85)$$

Thus $\mathbb{P}(f(\bar{u}) < \infty) = 1$, which proves the assertion. \square

4 Proof of the Main Theorem

In order to show the bound (46) and Thm. 1, we first need several technical propositions regarding the composition of two BV functions along a set of finite perimeter. We denote by $(E)^1$ and $(E)^0$ the measure-theoretic interior and exterior of a set E , see [2],

$$(E)^t := \{x \in \Omega \mid \lim_{\rho \searrow 0} \frac{|\mathcal{B}_\rho(x) \cap E|}{|\mathcal{B}_\rho(x)|} = t\}, \quad t \in [0, 1]. \quad (86)$$

Here $\mathcal{B}_\rho(x)$ denotes the ball with radius ρ centered in x , and $|A| := \mathcal{L}^d(A)$ the Lebesgue content of a set $A \subseteq \mathbb{R}^d$.

Proposition 3 *Let Ψ be positively homogeneous and convex, and satisfy the upper-boundedness condition (24). Then*

$$\Psi(\nu(z^1 - z^2)^\top) \leq \lambda_u \quad \forall z^1, z^2 \in \Delta_l. \quad (87)$$

Moreover, there exists a constant $C < \infty$ such that

$$\Psi(w) \leq C\|w\|_2 \quad \forall w \in W, \quad (88)$$

$$W := \{w = (w^1 | \dots | w^l) \in \mathbb{R}^{d \times l} \mid \sum_{i=1}^l w^i = 0\}. \quad (89)$$

Proof See appendix.

Proposition 4 *Let $E, F \subseteq \Omega^d$ be \mathcal{L}^d -measurable sets.*

Then

$$(E \cap F)^1 = (E)^1 \cap (F)^1. \quad (90)$$

Proof See appendix.

Proposition 5 *Let $u, v \in \text{BV}(\Omega, \Delta_l)$ and $E \subseteq \Omega$ such that $\text{Per}(E) < \infty$. Define*

$$w := u1_E + v1_{\Omega \setminus E}. \quad (91)$$

Then $w \in \text{BV}(\Omega, \Delta_l)^l$, and

$$Dw = Du \llcorner (E)^1 + Dv \llcorner (E)^0 + \nu_E (u_{\mathcal{F}E}^+ - v_{\mathcal{F}E}^-)^\top \mathcal{H}^{d-1} \llcorner (\mathcal{F}E \cap \Omega), \quad (92)$$

where $u_{\mathcal{F}E}^+$ and $v_{\mathcal{F}E}^-$ denote the one-sided approximate limits of u and v on the reduced boundary $\mathcal{F}E$, and ν_E is the generalized inner normal of E [2]. Moreover, for continuous, convex and positively homogeneous Ψ satisfying the upper-boundedness condition (24) and some Borel set $A \subseteq \Omega$,

$$\int_A d\Psi(Dw) \leq \int_{A \cap (E)^1} d\Psi(Du) + \int_{A \cap (E)^0} d\Psi(Dv) + \lambda_u \text{Per}(E). \quad (93)$$

Proof See appendix.

Proposition 6 *Let $u, v \in \text{BV}(\Omega, \Delta_l)$, $E \subseteq \Omega$ such that $\text{Per}(E) < \infty$, and*

$$u|_{(E)^1} = v|_{(E)^1} \quad \mathcal{L}^d\text{-a.e.} \quad (94)$$

Then $(Du) \llcorner (E)^1 = (Dv) \llcorner (E)^1$, and $\Psi(Du) \llcorner (E)^1 = \Psi(Dv) \llcorner (E)^1$. In particular,

$$\int_{(E)^1} d\Psi(Du) = \int_{(E)^1} d\Psi(Dv). \quad (95)$$

The result also holds when $(E)^1$ is replaced by $(E)^0$. Moreover, the condition (94) is equivalent to

$$u|_E = v|_E \quad \mathcal{L}^d\text{-a.e.} \quad (96)$$

Proof See appendix.

Remark 1 Note that taking the measure-theoretic interior $(E)^1$ is of central importance. The corollary does not hold when replacing the integral over $(E)^1$ with the integral over E , as can be seen from the example of the closed unit ball, i.e., $E = \mathcal{B}_1(0)$, $u = 1_E$ and $v \equiv 1$.

4.1 Proof of Theorem 1

In Sect. 3.2 we have shown that the rounding process induced by Alg. 1 is well-defined in the sense that it returns an integral solution $\bar{u}_\gamma \in \text{BV}(\Omega)^l$ almost surely. We now return to proving an upper bound for the expectation of $f(\bar{u})$ as in the approximate coarea formula (43). We first show that the expectation of the *linear part* (data term) of f is invariant under the rounding process.

Proposition 7 *The sequence (u^k) generated by Alg. 1 satisfies*

$$\mathbb{E}(\langle u^k, s \rangle) = \langle u, s \rangle \quad \forall k \in \mathbb{N}. \quad (97)$$

Proof In Alg. 1, instead of step 5 we consider the simpler update

$$u^k \leftarrow e^{i^k} 1_{\{u_{i^k}^{k-1} > \alpha^k\}} + u^{k-1} 1_{\{u_{i^k}^{k-1} \leq \alpha^k\}}. \quad (98)$$

This yields exactly the same sequence (u^k) , since if $u_{i^k}^{k-1}(x) > \alpha^k$ for any $\alpha^k \geq 0$, then either $x \in U^{k-1}$, or $u_{i^k}^{k-1}(x) = 1$. In both algorithms, points that are assigned a label e^{i^k} at some point in the process will never be assigned a *different* label at a later point. This is made explicit in Alg. 1 by keeping track of the set U^k of yet unassigned points. In contrast, using the step (98), a point may formally be assigned the same label multiple times.

Denote $\gamma' := (\gamma^1, \dots, \gamma^{k-1})$ and $u^{\gamma'} := u_\gamma^{k-1}$. We apply induction on k : For $k \geq 1$,

$$\begin{aligned} & \mathbb{E}_\gamma \langle u_\gamma^k, s \rangle \quad (99) \\ &= \mathbb{E}_{\gamma'} \frac{1}{l} \sum_{i=1}^l \int_0^1 \sum_{j=1}^l s_j \cdot \left(e^{i^k} 1_{\{u_i^{\gamma'} > \alpha\}} + u^{\gamma'} 1_{\{u_i^{\gamma'} \leq \alpha\}} \right)_j d\alpha \\ &= \mathbb{E}_{\gamma'} \frac{1}{l} \sum_{i=1}^l \int_0^1 \left(s_i \cdot 1_{\{u_i^{\gamma'} > \alpha\}} + u^{\gamma'} 1_{\{u_i^{\gamma'} \leq \alpha\}} \langle u^{\gamma'}, s \rangle \right) d\alpha \\ &= \mathbb{E}_{\gamma'} \frac{1}{l} \sum_{i=1}^l \int_0^1 \left(s_i \cdot 1_{\{u_i^{\gamma'} > \alpha\}} + \left(1 - 1_{\{u_i^{\gamma'} > \alpha\}} \right) \langle u^{\gamma'}, s \rangle \right) d\alpha. \quad (100) \end{aligned}$$

Now we take into account the property [2, Prop. 1.78], which is a direct consequence of Fubini's theorem, and also used in the proof of the thresholding theorem for the two-class case [7]:

$$\int_0^1 \int_\Omega s_i(x) \cdot 1_{\{u_i > \alpha\}}(x) dx d\alpha \quad (101)$$

$$= \int_\Omega s_i(x) u_i(x) dx = \langle u_i, s_i \rangle. \quad (102)$$

This leads to

$$\begin{aligned} & \mathbb{E}_\gamma \langle u_\gamma^k, s \rangle \\ &= \mathbb{E}_{\gamma'} \frac{1}{l} \sum_{i=1}^l \left(s_i u_i^{\gamma'} + \langle u^{\gamma'}, s \rangle - u_i^{\gamma'} \langle u^{\gamma'}, s \rangle \right) d\alpha \quad (103) \end{aligned}$$

and therefore, using $u^{\gamma'}(x) \in \Delta_l$,

$$\mathbb{E}_\gamma \langle u_\gamma^k, s \rangle = \mathbb{E}_{\gamma'} \langle u^{\gamma'}, s \rangle = \mathbb{E}_\gamma \langle u_\gamma^{k-1}, s \rangle. \quad (104)$$

Since $\langle u^0, s \rangle = \langle u, s \rangle$, the assertion follows by induction. \square

Remark 2 Prop. 7 shows that the data term is – in the mean – not affected by the probabilistic rounding process, i.e., it satisfies an *exact* coarea-like formula, even in the multiclass case.

Bounding the regularizer is more involved: For $\gamma^k = (i^k, \alpha^k)$, define

$$U_{\gamma^k} := \{x \in \Omega | u_{i^k}(x) \leq \alpha^k\}, \quad (105)$$

$$V_{\gamma^k} := (U_{\gamma^k})^1, \quad (106)$$

$$V^k := (U^k)^1. \quad (107)$$

As the measure-theoretic interior is invariant under \mathcal{L}^d -negligible modifications, given some fixed sequence γ the sequence (V^k) is invariant under \mathcal{L}^d -negligible modifications of $u = u^0$, i.e., it is uniquely defined when viewing u as an element of $L^1(\Omega)^l$. Some calculations yield

$$U^k = U_{\gamma^1} \cap \dots \cap U_{\gamma^k}, \quad k \geq 1, \quad (108)$$

$$U^{k-1} \setminus U^k = U_{\gamma^1} \cap \left((U_{\gamma^2} \cap \dots \cap U_{\gamma^{k-1}}) \setminus (U_{\gamma^2} \cap \dots \cap U_{\gamma^k}) \right), \quad k \geq 2. \quad (109)$$

From these observations and Prop. 4,

$$V^k = V_{\gamma^1} \cap \dots \cap V_{\gamma^k}, \quad k \geq 1, \quad (110)$$

$$V^{k-1} \setminus V^k = V_{\gamma^1} \cap \left((V_{\gamma^2} \cap \dots \cap V_{\gamma^{k-1}}) \setminus (V_{\gamma^2} \cap \dots \cap V_{\gamma^k}) \right), \quad k \geq 2, \quad (111)$$

$$\Omega \setminus V^k = \bigcup_{k'=1}^k (V^{k'-1} \setminus V^k), \quad k \geq 1. \quad (112)$$

Moreover, since V^k is the measure-theoretic interior of U^k , both sets are equal up to an \mathcal{L}^d -negligible set (cf. (174)).

We now prepare for an induction argument on the expectation of the regularizing term when restricted to the sets $V^{k-1} \setminus V^k$. The following proposition provides the initial step ($k = 1$).

Proposition 8 *Assume that Ψ satisfies the lower- and upper-boundedness conditions (23) and (24). Then*

$$\mathbb{E} \int_{V^0 \setminus V^1} d\Psi(D\bar{u}) \leq \frac{2}{l} \frac{\lambda_u}{\lambda_l} \int_\Omega d\Psi(Du). \quad (113)$$

Proof Denote $(i, \alpha) = \gamma^1$. Since $1_{U_{(i, \alpha)}} = 1_{V_{(i, \alpha)}}$ \mathcal{L}^d -a.e., we have

$$\bar{u}_\gamma = 1_{V_{(i, \alpha)}} e^i + 1_{\Omega \setminus V_{(i, \alpha)}} \bar{u}_\gamma \quad \mathcal{L}^d\text{-a.e.} \quad (114)$$

Therefore, since $V^0 = (U^0)^1 = (\Omega)^1 = \Omega$,

$$\begin{aligned} \int_{V^0 \setminus V^1} d\Psi(D\bar{u}_\gamma) &= \int_{\Omega \setminus V_{(i, \alpha)}} d\Psi(D\bar{u}_\gamma) \\ &= \int_{\Omega \setminus V_{(i, \alpha)}} d\Psi \left(D \left(1_{V_{(i, \alpha)}} e^i + 1_{\Omega \setminus V_{(i, \alpha)}} \bar{u}_\gamma \right) \right). \end{aligned} \quad (115)$$

Since $u \in \text{BV}(\Omega)^l$, we know that $\text{Per}(V_{(i, \alpha)}) < \infty$ holds for \mathcal{L}^1 -a.e. α and any i [2, Thm. 3.40]. Therefore we conclude from Prop. 5 that for \mathcal{L}^1 -a.e. α ,

$$\begin{aligned} \int_{\Omega \setminus V_{(i, \alpha)}} d\Psi(D\bar{u}_\gamma) &\leq \lambda_u \text{Per}(V_{(i, \alpha)}) + \\ &\int_{(\Omega \setminus V_{(i, \alpha)}) \cap (\Omega \setminus V_{(i, \alpha)})^1} d\Psi(De^i) + \\ &\int_{(\Omega \setminus V_{(i, \alpha)}) \cap (\Omega \setminus V_{(i, \alpha)})^0} d\Psi(D\bar{u}_\gamma). \end{aligned} \quad (116)$$

Both of the integrals are zero, since $De^i = 0$ and

$$(\Omega \setminus V_{(i, \alpha)})^0 = (V_{(i, \alpha)})^1 = V_{(i, \alpha)}, \quad (117)$$

therefore

$$\int_{\Omega \setminus V_{(i, \alpha)}} d\Psi(D\bar{u}_\gamma) \leq \lambda_u \text{Per}(V_{(i, \alpha)}). \quad (118)$$

Carrying the bound over to the expectation yields

$$\mathbb{E}_\gamma \int_{\Omega \setminus V_{(i, \alpha)}} d\Psi(D\bar{u}_\gamma) \leq \frac{1}{l} \sum_{i=1}^l \int_0^1 \lambda_u \text{Per}(V_{(i, \alpha)}) d\alpha.$$

Also, $\text{Per}(V_{(i, \alpha)}) = \text{Per}(U_{(i, \alpha)})$ since the perimeter is invariant under \mathcal{L}^d -negligible modifications. The assertion then follows using $V^0 = \Omega$, $V^1 = V_{(i, \alpha)}$ and the coarea formula:

$$\mathbb{E}_\gamma \int_{V^0 \setminus V^1} d\Psi(D\bar{u}_\gamma) \quad (119)$$

$$\leq \frac{1}{l} \sum_{i=1}^l \int_0^1 \lambda_u \text{Per}(U_{(i, \alpha)}) d\alpha \quad (120)$$

$$\stackrel{\text{coarea}}{=} \frac{\lambda_u}{l} \sum_{i=1}^l \text{TV}(u_i) = \frac{\lambda_u}{l} \int_\Omega \sum_{i=1}^l d\|Du_i\|_2 \quad (121)$$

$$\stackrel{(23)}{\leq} \frac{2\lambda_u}{l\lambda_l} \int_\Omega d\Psi(Du). \quad (122)$$

We now take care of the induction step for the regularizer bound.

Proposition 9 Let Ψ satisfy the upper-boundedness condition (24). Then, for any $k \geq 2$,

$$F := \mathbb{E} \int_{V^{k-1} \setminus V^k} d\Psi(D\bar{u}) \quad (123)$$

$$\leq \frac{(l-1)}{l} \mathbb{E} \int_{V^{k-2} \setminus V^{k-1}} d\Psi(D\bar{u}). \quad (124)$$

Proof Define the shifted sequence $\gamma' = (\gamma'^k)_{k=1}^\infty$ by $\gamma'^k := \gamma^{k+1}$, and let

$$W_{\gamma'} := V_{\gamma'}^{k-2} \setminus V_{\gamma'}^{k-1} \quad (125)$$

$$= (V_{\gamma^2} \cap \dots \cap V_{\gamma^{k-1}}) \setminus (V_{\gamma^2} \cap \dots \cap V_{\gamma^k}). \quad (126)$$

By Prop. 2 we may assume that, under the expectation, $\bar{u}_{\gamma'}$ exists and is an element of $\text{BV}(\Omega)^l$. We denote $\gamma^1 = (i, \alpha)$, then $V^{k-1} \setminus V^k = V_{(i, \alpha)} \cap W_{\gamma'}$ due to (111). For each pair (i, α) we denote by $((i, \alpha), \gamma')$ the sequence obtained by prepending (i, α) to the sequence γ' . Then

$$F = \frac{1}{l} \sum_{i=1}^l \int_0^1 \mathbb{E}_{\gamma'} \int_{V_{(i, \alpha)} \cap W_{\gamma'}} d\Psi(D\bar{u}_{((i, \alpha), \gamma')}) d\alpha. \quad (127)$$

Since in the first iteration of the algorithm no points in $U_{(i, \alpha)}$ are assigned a label, $\bar{u}_{((i, \alpha), \gamma')} = \bar{u}_{\gamma'}$ holds on $U_{(i, \alpha)}$, and therefore \mathcal{L}^d -a.e. on $V_{(i, \alpha)}$. Therefore we may apply Prop. 6 and substitute $D\bar{u}_{((i, \alpha), \gamma')}$ by $D\bar{u}_{\gamma'}$ in (127):

$$F = \frac{1}{l} \sum_{i=1}^l \int_0^1 \left(\mathbb{E}_{\gamma'} \int_{V_{(i, \alpha)} \cap W_{\gamma'}} d\Psi(D\bar{u}_{\gamma'}) \right) d\alpha \quad (128)$$

$$= \frac{1}{l} \sum_{i=1}^l \int_0^1 \left(\mathbb{E}_{\gamma'} \int_{W_{\gamma'}} 1_{V_{(i, \alpha)}} d\Psi(D\bar{u}_{\gamma'}) \right) d\alpha. \quad (129)$$

By definition of the measure-theoretic interior (86), the indicator function $1_{V_{(i, \alpha)}}$ is bounded from above by the density function $\Theta_{U_{(i, \alpha)}}$ of $U_{(i, \alpha)}$,

$$1_{V_{(i, \alpha)}}(x) \leq \Theta_{(i, \alpha)}(x) := \lim_{\delta \searrow 0} \frac{|\mathcal{B}_\delta(x) \cap U_{(i, \alpha)}|}{|\mathcal{B}_\delta(x)|}, \quad (130)$$

which exists \mathcal{H}^{d-1} -a.e. on Ω by [2, Prop. 3.61]. Therefore, denoting by $\mathcal{B}_\delta(\cdot)$ the mapping $x \in \Omega \mapsto \mathcal{B}_\delta(x)$,

$$F \leq \frac{1}{l} \sum_{i=1}^l \int_0^1 \mathbb{E}_{\gamma'} \int_{W_{\gamma'}} \lim_{\delta \searrow 0} \frac{|\mathcal{B}_\delta(\cdot) \cap U_{(i, \alpha)}|}{|\mathcal{B}_\delta(\cdot)|} d\Psi(D\bar{u}_{\gamma'}) d\alpha.$$

Rearranging the integrals and the limit, which can be justified by $\text{TV}(\bar{u}_{\gamma'}) < \infty$ almost surely and dominated convergence using (24), we get

$$\begin{aligned} F &\leq \frac{1}{l} \mathbb{E}_{\gamma'} \lim_{\delta \searrow 0} \int_{W_{\gamma'}} \sum_{i=1}^l \int_0^1 \frac{|\mathcal{B}_\delta(\cdot) \cap U_{(i, \alpha)}|}{|\mathcal{B}_\delta(\cdot)|} d\alpha d\Psi(D\bar{u}_{\gamma'}) \\ &= \frac{1}{l} \mathbb{E}_{\gamma'} \lim_{\delta \searrow 0} \int_{W_{\gamma'}} \frac{1}{|\mathcal{B}_\delta(\cdot)|} \cdot \\ &\quad \left(\sum_{i=1}^l \int_0^1 \int_{\mathcal{B}_\delta(\cdot)} 1_{\{u_i(y) \leq \alpha\}} dy d\alpha \right) d\Psi(D\bar{u}_{\gamma'}). \end{aligned} \quad (131)$$

We again apply [2, Prop. 1.78] to the two innermost integrals (alternatively, use Fubini's theorem), which leads to

$$F \leq \frac{1}{l} \mathbb{E}_{\gamma'} \lim_{\delta \searrow 0} \int_{W_{\gamma'}} \frac{1}{|\mathcal{B}_\delta(\cdot)|} . \quad (132)$$

$$\left(\sum_{i=1}^l \int_{\mathcal{B}_\delta(\cdot)} (1 - u_i(y)) dy \right) d\Psi(D\bar{u}_{\gamma'}) . \quad (133)$$

Using the fact that $u(y) \in \Delta_l$, this collapses according to

$$\begin{aligned} F &\leq \frac{1}{l} \mathbb{E}_{\gamma'} \lim_{\delta \searrow 0} \int_{W_{\gamma'}} \frac{1}{|\mathcal{B}_\delta(\cdot)|} \left(\int_{\mathcal{B}_\delta(\cdot)} (l-1) dy \right) d\Psi(D\bar{u}_{\gamma'}) \\ &= \frac{1}{l} \mathbb{E}_{\gamma'} \lim_{\delta \searrow 0} \int_{W_{\gamma'}} (l-1) d\Psi(D\bar{u}_{\gamma'}) \end{aligned} \quad (134)$$

$$= \frac{l-1}{l} \mathbb{E}_{\gamma'} \int_{W_{\gamma'}} d\Psi(D\bar{u}_{\gamma'}) \quad (135)$$

$$= \frac{l-1}{l} \mathbb{E}_{\gamma'} \int_{V_{\gamma'}^{k-2} \setminus V_{\gamma'}^{k-1}} d\Psi(D\bar{u}_{\gamma'}) . \quad (136)$$

Reverting the index shift and using $\bar{u}_{\gamma'} = \bar{u}_\gamma$ concludes the proof:

$$F \leq \frac{l-1}{l} \mathbb{E}_\gamma \int_{V_\gamma^{k-1} \setminus V_\gamma^k} d\Psi(D\bar{u}_\gamma) . \quad (137)$$

We are now ready to prove the main result, Thm. 1, as stated in the introduction.

Proof (Theorem 1) The fact that the algorithm provides $\bar{u} \in \mathcal{C}_\mathcal{E}$ almost surely follows from Thm. 2. Therefore there almost surely exists $k' := k'(\gamma) \geq 1$ such that $U^{k'} = \emptyset$ and $\bar{u}_\gamma = u_\gamma^{k'}$. On one hand, this implies

$$\langle \bar{u}_\gamma, s \rangle = \langle u_\gamma^{k'}, s \rangle = \lim_{k \rightarrow \infty} \langle u_\gamma^k, s \rangle \quad (138)$$

almost surely. On the other hand, $V^{k'} = (U^{k'})^1 = \emptyset$ and therefore

$$\bigcup_{k=1}^{k'} V^{k-1} \setminus V^k \stackrel{(*)}{=} \Omega \setminus V^{k'} = \Omega \quad (139)$$

almost surely. The equality $(*)$ can be shown by induction: For the base case $k' = 1$, we have $V^0 = (U^0)^1 = (\Omega)^1 = \Omega$, since Ω is the open unit box. For $k' \geq 2$,

$$\bigcup_{k=1}^{k'} V^{k-1} \setminus V^k \quad (140)$$

$$= \left(V^{k'-1} \setminus V^{k'} \right) \cup \bigcup_{k=1}^{k'-1} (V^{k-1} \setminus V^k) \quad (141)$$

$$= \left(V^{k'-1} \setminus V^{k'} \right) \cup \left(\Omega \setminus V^{k'-1} \right) \quad (142)$$

$$V^{k'-1} \subseteq \Omega \quad \Omega \setminus V^{k'-1} . \quad (143)$$

almost surely (cf. (112)). From (138) and (139) we obtain

$$\begin{aligned} \mathbb{E}_\gamma f(\bar{u}_\gamma) &= \mathbb{E}_\gamma \left(\lim_{k \rightarrow \infty} \langle u_\gamma^k, s \rangle \right) + \\ &\quad \mathbb{E}_\gamma \left(\sum_{k=1}^{\infty} \int_{V^{k-1} \setminus V^k} d\Psi(D\bar{u}_\gamma) \right) \end{aligned} \quad (144)$$

$$= \lim_{k \rightarrow \infty} (\mathbb{E}_\gamma \langle u_\gamma^k, s \rangle) + \sum_{k=1}^{\infty} \mathbb{E}_\gamma \int_{V^{k-1} \setminus V^k} d\Psi(D\bar{u}_\gamma) \quad (145)$$

The first term in (145) is equal to $\langle u, s \rangle$ due to Prop. 7. An induction argument using Prop. 8 and Prop. 9 shows that the second term can be bounded according to

$$\sum_{k=1}^{\infty} \mathbb{E}_\gamma \int_{V^{k-1} \setminus V^k} d\Psi(D\bar{u}_\gamma) \quad (146)$$

$$\leq \sum_{k=1}^{\infty} \left(\frac{l-1}{l} \right)^{k-1} \frac{2}{l} \frac{\lambda_u}{\lambda_l} \int_{\Omega} d\Psi(Du) \quad (147)$$

$$= 2 \frac{\lambda_u}{\lambda_l} \int_{\Omega} d\Psi(Du) , \quad (148)$$

therefore

$$\mathbb{E}_\gamma f(\bar{u}_\gamma) \leq \langle u, s \rangle + 2 \frac{\lambda_u}{\lambda_l} \int_{\Omega} d\Psi(Du) . \quad (149)$$

Since $s \geq 0$ and $\lambda_u \geq \lambda_l$, and therefore the linear term is bounded by $\langle u, s \rangle \leq 2(\lambda_u/\lambda_l)\langle u, s \rangle$, this proves the assertion. Swapping the integral and limit in (145) can be justified retrospectively by the dominated convergence theorem, using $0 \leq \langle u, s \rangle \leq \infty$ and $\int_{\Omega} d\Psi(Du) < \infty$ due to the upper-boundedness condition and Prop. 3. \square

Corollary 1 (see introduction) follows immediately using $f(u^*) \leq f(u_\mathcal{E}^*)$, cf. (44). We have demonstrated that the proposed approach allows to recover, from the solution u^* of the convex *relaxed* problem (13), an approximate *integral* solution \bar{u}^* of the nonconvex *original* problem (1) with an upper bound on the objective.

For the specific case $\Psi = \Psi_d$, we have

Proposition 10 *Let $d : \mathcal{I}^2 \rightarrow \mathbb{R}_{\geq 0}$ be a metric and $\Psi = \Psi_d$. Then one may set*

$$\lambda_u = \max_{i,j \in \{1, \dots, l\}} d(i,j) \text{ and } \lambda_l = \min_{i \neq j} d(i,j).$$

Proof From the remarks in the introduction we obtain (cf. [19])

$$\Psi_d(\nu(e^i - e^j)) = d(i,j),$$

which shows the upper bound. For the lower bound, set $c := \min_{i \neq j} d(i,j)$, $v'^i := \frac{c}{2} \frac{w^i}{\|w^i\|_2}$ and $v := v'(I - \frac{1}{l}ee^\top)$.

Then $v \in \mathcal{D}_{\text{loc}}^d$, since $\|v^i - v^j\|_2 = \|v'^i - v'^j\|_2 \leq c$ and $ve = v'(I - \frac{1}{l}ee^\top)e = 0$. Therefore, for $w \in \mathbb{R}^{d \times l}$ satisfying $we = 0$,

$$\Psi_d(w) \geq \langle w, v \rangle = \langle w, v' \rangle \quad (150)$$

$$= \sum_{i=1}^l \langle w^i, \frac{c}{2} \frac{w^i}{\|w^i\|_2} \rangle = \frac{c}{2} \sum_{i=1}^l \|w^i\|_2, \quad (151)$$

proving the lower bound.

Finally, for Ψ_d we obtain the factor

$$2 \frac{\lambda_u}{\lambda_l} = 2 \frac{\max_{i,j} d(i,j)}{\min_{i \neq j} d(i,j)}, \quad (152)$$

determining the optimality bound, as claimed in the introduction (28). The bound in (27) is the same as the known bounds for finite-dimensional metric labeling [12] and α -expansion [4], however it extends these results to problems on continuous domains for a broad class of regularizers.

5 Conclusion

In this work we considered a method for recovering approximate solutions of image partitioning problems from solutions of a convex relaxation. We proposed a probabilistic rounding method motivated by the finite-dimensional framework, and showed that it is possible to obtain *a priori* bounds on the optimality of the integral solution obtained by rounding a solution of the convex relaxation.

The obtained bounds are compatible with known bounds for the finite-dimensional setting. However, to our knowledge, this is the first fully convex approach that is both formulated in the spatially continuous setting and provides a true *a priori* bound. We showed that the approach can also be interpreted as an approximate variant of the coarea formula.

While the results apply to a quite general class of regularizers, they are formulated for the homogeneous case. Non-homogeneous regularizers constitute an interesting direction for future work. In particular, such regularizers naturally occur when applying convex relaxation techniques [1, 25] in order to solve nonconvex variational problems.

With the increasing computational power, such techniques have become quite popular recently. For problems where the convexity is confined to the data term, they permit to find a global minimizer. A proper extension of the results outlined in this work may provide a way to find good approximate solutions of problems where also the *regularizer* is nonconvex.

6 Appendix

Proof (Prop. 3) In order to prove the first assertion (87), note that the mapping $w \mapsto \Psi(vw^\top)$ is convex, therefore it must assume its maximum on the polytope $\Delta_l - \Delta_l := \{z^1 - z^2 | z^1, z^2 \in \Delta_l\}$ in a vertex of the polytope. Since the polytope $\Delta_l - \Delta_l$ is the difference of two polytopes, its vertex set is at most the difference of their vertex sets, $V := \{e^i - e^j | i, j \in \{1, \dots, l\}\}$. On this set, the bound $\Psi(vw^\top) \leq \lambda_u$ holds for $w \in V$ due to the upper-boundedness condition (24), which proves (87).

The second equality (89) follows from the fact that $G := \{b^{ik} := e^k(e^i - e^{i+1})^\top | 1 \leq i \leq l-1, 1 \leq k \leq d\}$ is a basis of the linear subspace W , satisfying $\Psi(b^{ik}) \leq \lambda_u$, and Ψ is positively homogeneous and convex, and thus subadditive. Specifically, there is a linear transform $T : W \rightarrow \mathbb{R}^{d \times (l-1)}$ such that $w = \sum_{i,k} b^{ik} \alpha_{ik}$ for $\alpha = T(w)$. Then

$$\Psi(w) = \Psi \left(\sum_{i,k} b^{ik} \alpha_{ik} \right) \quad (153)$$

$$\leq \Psi \left(\sum_{ik} |\alpha_{ik}| \operatorname{sgn}(\alpha_{ik}) b^{ik} \right) \quad (154)$$

$$\leq \sum_{ik} |\alpha_{ik}| \Psi(\operatorname{sgn}(\alpha_{ik}) b^{ik}). \quad (155)$$

Since (24) provides $\Psi(\pm b^{ik}) \leq \lambda_u$, we obtain

$$\Psi(w) \leq \lambda_u \sum_{ik} |\alpha_{ik}| \leq \lambda_u \|T\| \|w\|_2 \quad (156)$$

for some suitable operator norm $\|\cdot\|$ and any $w \in W$.

Proof (Prop. 4) We prove mutual inclusion:

" \subseteq ": From the definition of the measure-theoretic interior,

$$x \in (E \cap F)^1 \Rightarrow \lim_{\delta \searrow 0} \frac{|\mathcal{B}_\delta(x) \cap E \cap F|}{|\mathcal{B}_\delta(x)|} = 1. \quad (157)$$

Since $|\mathcal{B}_\delta(x)| \geq |\mathcal{B}_\delta(x) \cap E| \geq |\mathcal{B}_\delta(x) \cap E \cap F|$ (and vice versa for $|\mathcal{B}_\delta(x) \cap F|$), it follows by the "sandwich" criterion that both $\lim_{\delta \searrow 0} |\mathcal{B}_\delta(x) \cap E|/|\mathcal{B}_\delta(x)|$ and $\lim_{\delta \searrow 0} |\mathcal{B}_\delta(x) \cap F|/|\mathcal{B}_\delta(x)|$ exist and are equal to 1, which shows $x \in E^1 \cap F^1$.

" \supseteq ": Assume that $x \in E^1 \cap F^1$. Then

$$1 \geq \limsup_{\delta \searrow 0} \frac{|\mathcal{B}_\delta(x) \cap E \cap F|}{|\mathcal{B}_\delta(x)|} \quad (158)$$

$$\geq \liminf_{\delta \searrow 0} \frac{|\mathcal{B}_\delta(x) \cap E \cap F|}{|\mathcal{B}_\delta(x)|} \quad (159)$$

$$= \liminf_{\delta \searrow 0} \frac{|\mathcal{B}_\delta(x) \cap E| + |\mathcal{B}_\delta(x) \cap F| - |\mathcal{B}_\delta(x) \cap (E \cup F)|}{|\mathcal{B}_\delta(x)|}.$$

We obtain equality,

$$1 \geq \liminf_{\delta \searrow 0} \frac{|\mathcal{B}_\delta(x) \cap E \cap F|}{|\mathcal{B}_\delta(x)|} \quad (160)$$

$$\geq \liminf_{\delta \searrow 0} \frac{|\mathcal{B}_\delta(x) \cap E|}{|\mathcal{B}_\delta(x)|} + \liminf_{\delta \searrow 0} \frac{|\mathcal{B}_\delta(x) \cap F|}{|\mathcal{B}_\delta(x)|} + \liminf_{\delta \searrow 0} \left(-\frac{|\mathcal{B}_\delta \cap (E \cup F)|}{|\mathcal{B}_\delta(x)|} \right) \quad (161)$$

$$= 2 - \underbrace{\limsup_{\delta \searrow 0} \frac{|\mathcal{B}_\delta \cap (E \cup F)|}{|\mathcal{B}_\delta(x)|}}_{\leq 1} \geq 1, \quad (162)$$

from which we conclude that

$$\limsup_{\delta \searrow 0} \frac{|\mathcal{B}_\delta(x) \cap E \cap F|}{|\mathcal{B}_\delta(x)|} = \liminf_{\delta \searrow 0} \frac{|\mathcal{B}_\delta(x) \cap E \cap F|}{|\mathcal{B}_\delta(x)|} = 1,$$

i.e., $x \in (E \cap F)^1$.

Proof (Prop. 5) First note that

$$\int_{\mathcal{F}E \cap \Omega} \|w_{\mathcal{F}E}^+ - w_{\mathcal{F}E}^-\|_2 d\mathcal{H}^{d-1} \quad (163)$$

$$\leq \sup\{\|w_{\mathcal{F}E}^+(x) - w_{\mathcal{F}E}^-(x)\|_2 \mid x \in \mathcal{F}E \cap \Omega\} \cdot \mathcal{H}^{d-1}(\mathcal{F}E \cap \Omega) \quad (164)$$

$$\stackrel{(*)}{\leq} \sup\{\|w(x) - w(y)\|_2 \mid x, y \in \Omega\} \cdot \text{TV}(1_E) \quad (165)$$

$$\stackrel{(**)}{\leq} \sqrt{2} \text{TV}(1_E) \quad (166)$$

$$= \sqrt{2} \text{Per}(E) < \infty. \quad (167)$$

The inequality $(*)$ is a consequence of the definition of $w_{\mathcal{F}E}^\pm$ and [2, Thm. 3.59], and $(**)$ follows directly from $w(x), w(y) \in \Delta_l$. The upper bound (167) permits applying [2, Thm. 3.84] on w , which provides $w \in \text{BV}(\Omega)^l$ and (92). Due to [2, Prop. 3.61], the sets $(E)^0, (E)^1$ and $\mathcal{F}E$ form a (pairwise disjoint) partition of Ω , up to an \mathcal{H}^{d-1} -zero set. Moreover, since $\Psi(Du) \ll |Du| \ll \mathcal{H}^{d-1}$ by construction, we have, for some Borel set A ,

$$\int_A \Psi(Dw) \quad (168)$$

$$= \int_{A \cap (E)^1} d\Psi(Dw) + \int_{A \cap (E)^0} d\Psi(Dw) + \quad (169)$$

$$\int_{A \cap \mathcal{F}E \cap \Omega} \Psi\left(\nu_E (w_{\mathcal{F}E}^+(x) - w_{\mathcal{F}E}^-(x))^\top\right) d\mathcal{H}^{d-1} \quad (170)$$

$$\stackrel{(**)}{\leq} \int_{A \cap (E)^1} d\Psi(Dw) + \int_{A \cap (E)^0} d\Psi(Dw) + \quad (170)$$

$$\int_{A \cap (E)^1} \lambda_u d\mathcal{H}^{d-1} \quad (171)$$

$$\stackrel{(167)}{\leq} \int_{A \cap (E)^1} d\Psi(Dw) + \int_{A \cap (E)^0} d\Psi(Dw) + \lambda_u \text{Per}(E).$$

The inequality $(**)$ holds due to the upper boundedness and Prop. 3. From [2, Prop. 2.37] we obtain that Ψ is

additive on mutually singular Radon measures μ, ν , i.e., if $|\mu| \perp |\nu|$, then

$$\int_B d\Psi(\mu + \nu) = \int_B d\Psi(\mu) + \int_B d\Psi(\nu) \quad (172)$$

for any Borel set $B \subseteq \Omega$. Substituting Dw in (171) according to (92) and using the fact that the three measures that form Dw in (92) are mutually singular, the additivity property (172) leads to the remaining assertion,

$$\int_A d\Psi(Dw) \leq \int_{A \cap (E)^1} d\Psi(Du) + \int_{A \cap (E)^0} d\Psi(Dv) + \lambda_u \text{Per}(E). \quad (173)$$

Proof (Prop. 6) We first show (96). It suffices to show that

$$\{x \in (E)^1 \Leftrightarrow x \in E\} \quad \text{for } \mathcal{L}^d\text{-a.e. } x \in \Omega. \quad (174)$$

This can be seen by considering the precise representative $\widetilde{1}_E$ of 1_E [2, Def. 3.63]: Starting with the definition,

$$x \in (E)^1 \Leftrightarrow \lim_{\delta \searrow 0} \frac{|E \cap \mathcal{B}_\delta(x)|}{|\mathcal{B}_\delta(x)|} = 1, \quad (175)$$

the fact that $\lim_{\delta \searrow 0} \frac{|\Omega \cap \mathcal{B}_\delta(x)|}{|\mathcal{B}_\delta(x)|} = 1$ implies

$$x \in (E)^1 \Leftrightarrow \lim_{\delta \searrow 0} \frac{|(\Omega \setminus E) \cap \mathcal{B}_\delta(x)|}{|\mathcal{B}_\delta(x)|} = 0 \quad (176)$$

$$\Leftrightarrow \lim_{\delta \searrow 0} \frac{1}{|\mathcal{B}_\delta(x)|} \int_{\mathcal{B}_\delta(x)} |1_E - 1| dy = 0 \quad (177)$$

$$\Leftrightarrow \widetilde{1}_E(x) = 1. \quad (178)$$

Substituting E by $\Omega \setminus E$, the same equivalence shows that $x \in (E)^0 \Leftrightarrow \widetilde{1}_{\Omega \setminus E}(x) = 1 \Leftrightarrow \widetilde{1}_E(x) = 0$. As $\mathcal{L}^d(\Omega \setminus ((E)^0 \cup (E)^1)) = 0$, this shows that $1_{E^1} = \widetilde{1}_E$ \mathcal{L}^d -a.e. Using the fact that $\widetilde{1}_E = 1_E$ [2, Prop. 3.64], we conclude that $1_{(E)^1} = 1_E$ \mathcal{L}^d -a.e., which proves (174) and therefore the assertion (96).

Since the measure-theoretic interior $(E)^1$ is defined over \mathcal{L}^d -integrals, it is invariant under \mathcal{L}^d -negligible modifications of E . Together with (174) this implies

$$((E)^1)^1 = (E)^1, \quad \mathcal{F}(E)^1 = \mathcal{F}E, \quad ((E)^1)^0 = (E)^0. \quad (179)$$

To show the relation $(Du)_\perp(E)^1 = (Dv)_\perp(E)^1$, consider

$$Du_\perp(E)^1 = D(1_{\Omega \setminus (E)^1} u + 1_{(E)^1} u)_\perp(E)^1 \quad (180)$$

$$\stackrel{(*)}{=} D(1_{\Omega \setminus (E)^1} u + 1_{(E)^1} v)_\perp(E)^1. \quad (181)$$

The equality $(*)$ holds due to the assumption (94), and due to the fact that $Df = Dg$ if $f = g$ \mathcal{L}^d -a.e. (see,

e.g., [2, Prop. 3.2]). We continue from (181) via

$$Du_{\mathbb{L}}(E)^1 \quad (182)$$

$$\stackrel{\text{Prop. 5}}{=} \{Du_{\mathbb{L}}((E)^1)^0 + Dv_{\mathbb{L}}((E)^1)^1 + \quad (183)$$

$$\nu_{(E)^1} (u_{\mathcal{F}E^1}^+ - v_{\mathcal{F}E^1}^-)^\top \mathcal{H}^{d-1} \llcorner (\mathcal{F}(E)^1 \cap \Omega)\} \llcorner (E)^1$$

$$\stackrel{(179)}{=} (Du_{\mathbb{L}}(E)^0 + Dv_{\mathbb{L}}(E)^1) \llcorner (E)^1 + \quad (184)$$

$$(\nu_{(E)^1} (u_{\mathcal{F}E^1}^+ - v_{\mathcal{F}E^1}^-)^\top \mathcal{H}^{d-1} \llcorner (\mathcal{F}E \cap \Omega)) \llcorner (E)^1$$

$$= Du_{\mathbb{L}}((E)^0 \cap (E)^1) + Dv_{\mathbb{L}}((E)^1 \cap (E)^1) + \quad (185)$$

$$\nu_{(E)^1} (u_{\mathcal{F}E^1}^+ - v_{\mathcal{F}E^1}^-)^\top \mathcal{H}^{d-1} \llcorner (\mathcal{F}E \cap \Omega \cap (E)^1)$$

$$= Dv_{\mathbb{L}}(E)^1. \quad (186)$$

Therefore $Du_{\mathbb{L}}(E)^1 = Dv_{\mathbb{L}}(E)^1$. Then,

$$\Psi(Du) \llcorner (E)^1 = \Psi(Du_{\mathbb{L}}(E)^1 + Du_{\mathbb{L}}(\Omega \setminus (E)^1)) \llcorner (E)^1 \quad (187)$$

$$\stackrel{(*)}{=} \Psi(Du_{\mathbb{L}}(E)^1) \llcorner (E)^1 + \Psi(Du_{\mathbb{L}}(\Omega \setminus (E)^1)) \llcorner (E)^1. \quad (188)$$

In the equality $(*)$ we used the additivity of Ψ on mutually singular Radon measures [2, Prop. 2.37]. By definition of the total variation, $|\mu \llcorner A| = |\mu| \llcorner A$ holds for any measure μ , therefore $|Du_{\mathbb{L}}(\Omega \setminus (E)^1)| = |Du| \llcorner (\Omega \setminus (E)^1)$ and $|Du_{\mathbb{L}}(\Omega \setminus (E)^1)|((E)^1) = 0$, which together with (again by definition) $\Psi(\mu) \ll |\mu|$ implies that the second term in (188) vanishes. Since all observations equally hold for v instead of u , we conclude

$$\Psi(Du) \llcorner (E)^1 = \Psi(Du_{\mathbb{L}}(E)^1) \llcorner (E)^1 \quad (189)$$

$$\stackrel{(186)}{=} \Psi(Dv_{\mathbb{L}}(E)^1) \llcorner (E)^1 \quad (190)$$

$$= \Psi(Dv) \llcorner (E)^1. \quad (191)$$

Equation (95) follows immediately.

References

- Alberti, G., Bouchitté, G., Dal Maso, G.: The calibration method for the Mumford-Shah functional and free-discontinuity problems. *Calc. Var. Part. Diff. Eq.* **16**(3), 299–333 (2003)
- Ambrosio, L., Fusco, N., Pallara, D.: Functions of Bounded Variation and Free Discontinuity Problems. Clarendon Press (2000)
- Bae, E., Yuan, J., Tai, X.C.: Global minimization for continuous multiphase partitioning problems using a dual approach. *Int. J. Comp. Vis.* **92**, 112–129 (2011)
- Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. *Patt. Anal. Mach. Intell.* **23**(11), 1222–1239 (2001)
- Chambolle, A., Cremers, D., Pock, T.: A convex approach for computing minimal partitions. *Tech. Rep.* 649, Ecole Polytechnique CMAP (2008)
- Chambolle, A., Darbon, J.: On total variation minimization and surface evolution using parametric maximum flows. *Int. J. Comp. Vis.* **84**, 288–307 (2009)
- Chan, T.F., Esedoglu, S., Nikolova, M.: Algorithms for finding global minimizers of image segmentation and denoising models. *J. Appl. Math.* **66**(5), 1632–1648 (2006)
- Darbon, J., Sigelle, M.: Image restoration with discrete constrained total variation part I: Fast and exact optimization. *J. Math. Imaging Vis.* **26**(3), 261–276 (2006)
- Darbon, J., Sigelle, M.: Image restoration with discrete constrained total variation part II: Levelable functions, convex priors and non-convex cases. *J. Math. Imaging Vis.* **26**(3), 277–291 (2006)
- Delaunoy, A., Fundana, K., Prados, E., Heyden, A.: Convex multi-region segmentation on manifolds. In: *Int. Conf. Comp. Vis.* (2009)
- Goldstein, T., Bresson, X., Osher, S.: Global minimization of Markov random field with applications to optical flow. *CAM Report 09-77*, UCLA (2009)
- Kleinberg, J.M., Tardos, E.: Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields. In: *Found. Comp. Sci.*, pp. 14–23 (1999)
- Klodt, M., Schoenemann, T., Kolev, K., Schikora, M., Cremers, D.: An experimental comparison of discrete and continuous shape optimization methods. In: *Europ. Conf. Comp. Vis. Marseille, France* (2008)
- Kolev, K., Klodt, M., Brox, T., Cremers, D.: Continuous global optimization in multiview 3d reconstruction. *Int. J. Comp. Vis.* **84**(1) (2009)
- Komodakis, N., Tziritas, G.: Approximate labeling via graph cuts based on linear programming. *Patt. Anal. Mach. Intell.* **29**(8), 1436–1453 (2007)
- Lellmann, J., Becker, F., Schnörr, C.: Convex optimization for multi-class image labeling with a novel family of total variation based regularizers. In: *Int. Conf. Comp. Vis.* (2009)
- Lellmann, J., Kappes, J., Yuan, J., Becker, F., Schnörr, C.: Convex multi-class image labeling by simplex-constrained total variation. In: *Scale Space and Var. Meth., LNCS*, vol. 5567, pp. 150–162 (2009)
- Lellmann, J., Lenzen, F., Schnörr, C.: Optimality bounds for a variational relaxation of the image partitioning problem. In: *Energy Min. Meth. Comp. Vis. Patt. Recogn.* (2011)
- Lellmann, J., Schnörr, C.: Continuous multiclass labeling approaches and algorithms. *SIAM J. Imaging Sci.* p. In press. (2011)
- Lysaker, M., Tai, X.C.: Iterative image restoration combining total variation minimization and a second-order functional. *Int. J. Comp. Vis.* **66**(1), 5–18 (2006)
- Øksendal, B.: Stochastic Differential Equations, 6th edn. Springer (2003)
- Olsson, C.: Global optimization in computer vision: Convexity, cuts and approximation algorithms. Ph.D. thesis, Lund Univ. (2009)
- Olsson, C., Byrød, M., Overgaard, N.C., Kahl, F.: Extending continuous cuts: Anisotropic metrics and expansion moves. In: *Int. Conf. Comp. Vis.* (2009)
- Paragios, N., Chen, Y., Faugeras, O. (eds.): The Handbook of Mathematical Models in Computer Vision. Springer (2006)
- Pock, T., Cremers, D., Bischof, H., Chambolle, A.: Global solutions of variational models with convex regularization. *J. Imaging Sci.* **3**(4), 1122–1145 (2010)
- Strandmark, P., Kahl, F., Overgaard, N.C.: Optimizing parametric total variation models. In: *Int. Conf. Comp. Vis.* (2009)
- Trobin, W., Pock, T., Cremers, D., Bischof, H.: Continuous energy minimization by repeated binary fusion. In: *Europ. Conf. Comp. Vis.*, vol. 4, pp. 667–690 (2008)

- 28. Yuan, J., Bae, E., Tai, X.C., Boykov, Y.: A continuous max-flow approach to potts model. In: Europ. Conf. Comp. Vis., pp. 379–392 (2010)
- 29. Zach, C., Gallup, D., Frahm, J.M., Niethammer, M.: Fast global labeling for real-time stereo using multiple plane sweeps. In: Vis. Mod. Vis (2008)